

Internet **Electronic** Journal of **Molecular Design**

August 2008, Volume 7, Number 8, Pages 161–185

Editor: Ovidiu Ivanciuc

2D–QSAR Autocorrelation Study on Selective COX–2 Inhibitors

Safia Taïri–Kellou,¹ Souhila Bouaziz–Terrachet,¹ Boubekour Maouche,¹
and Gilles Moreau²

¹ Laboratoire de Physico–chimie Théorique et Chimie Informatique, USTHB, Alger, Algérie
² 30 avenue Jean Jaurès, 94220 Charenton le Pont, France

Received: June 10, 2007; Revised: February 7, 2008; Accepted: June 4, 2008; Published: August 31, 2008

Citation of the article:

S. Taïri–Kellou, S. Bouaziz–Terrachet, B. Maouche, and G. Moreau, 2D–QSAR Autocorrelation Study on Selective COX–2 Inhibitors, *Internet Electron. J. Mol. Des.* **2008**, 7, 161–185, <http://www.biochempress.com>.

2D–QSAR Autocorrelation Study on Selective COX–2 Inhibitors

Safia Taïri–Kellou,^{1,*} Souhila Bouaziz–Terrachet,¹ Boubekour Maouche,¹
and Gilles Moreau²

¹ Laboratoire de Physico–chimie Théorique et Chimie Informatique, USTHB, Alger, Algeria
² 30 avenue Jean Jaurès, 94220 Charenton le Pont, France

Received: June 10, 2007; Revised: February 7, 2008; Accepted: June 4, 2008; Published: August 31, 2008

Internet Electron. J. Mol. Des. 2008, 7 (8), 161–185

Abstract

Motivation. A database of 185 COX–2 inhibitors including traditional NSAIDs was used to derive regression and classification models starting from the 2D autocorrelation descriptors. For that, we have described an original application of autocorrelation vectors for encoding the chemical information derived from 2D–structures.

Method. Thus, a separate autocorrelation vector was computed for each of the following atomic properties: Pauling electronegativity, van der Waals volume, indicator for saturated/unsaturated, logP contribution, polar surface area, indicator for hydrogen–bond donor, indicator for hydrogen–bond acceptor and partial charge, and the resulting set was reduced into a smaller number of variables using PCA.

Results. The robustness and predictive ability of the obtained models were evaluated by selecting chemical classes similar to the training set with activities ranging from low to high and the results of the models were compared.

Conclusions. The data generated from the present study should be useful for predictive purposes, as an *in silico* filter, to screen large structural database for new potent COX–2 inhibitor.

Keywords. 2D–QSAR; autocorrelation vectors; cyclooxygenase–2 inhibitors; principal component analysis; PCA; multiple linear regression; MLR; linear discriminant analysis; LDA.

Abbreviations and notations

2D, two–dimensional	NSAIDs, nonsteroidal anti–inflammatory drugs
3D, three–dimensional	PCA, principal component analysis
COX, cyclooxygenase	PCs, principal components
FP, false positive	QSAR, quantitative structure–activity relationships
FN, false negative	SAR, structure–activity relationships
LDA, linear discriminant analysis	TN, true negative
MLR, multiple linear regression	TP, true positive

* Correspondence author; phone: 00–213–73 41 00 15; fax: 00–213 21 81 04 34; E–mail: s_kellou@yahoo.fr.

1 INTRODUCTION

The classical non-steroidal anti-inflammatory drugs (NSAIDs) are widely used to alleviate the symptoms of inflammatory diseases. The principal pharmacological effect of NSAIDs is their ability to inhibit prostaglandin synthesis by blocking cyclooxygenase (COX) activity [1,2]. Two isoforms of the COX enzyme, COX-1 and COX-2, with their functional roles in the maintenance of normal physiological functions including gastric cytoprotection and induction of the high-level production of prostaglandins (PGs) that results in inflammation, respectively, were identified early in the past decade [3,4]. Thus, an NSAID that inhibits COX-2 while sparing COX-1 has the potential to be anti-inflammatory yet nontoxic in the gastrointestinal tract. This hypothesis has generated a great deal of interest in this field and various laboratories are aggressively pursuing this objective. To speed up drug development for safer NSAIDs, the discovery and characterization of the second isoform (COX-2) has taken a new turn, to have exclusive control over inflammatory processes, leaving other tissues unharmed even during chronic treatment.

Recent epidemiological and experimental data suggest that selective inhibition of COX-2 could also be an important strategy for preventing or treating number of diseases such as cancer, Alzheimer's disease, cardiovascular and blood clotting disorders [5]. In the last several years, extensive libraries of selective COX-2 inhibitors have been developed. However, some COX-2 selective inhibitors on the market, namely current generation drugs, have been associated with adverse cardiovascular side effects, leading to recent restrictions on their marketing. Rofecoxib (Merck) [6] has been withdrawn from the market because of its increased risk of cardiovascular problems [7]. Thus, there is a need for designing new selective COX-2 inhibitors. The physiology and pathophysiology of COX continue, therefore, to be a matter of great interest [8-12], and selective inhibitors of the COX-2 enzyme have received a great deal of attention in the recent literature as superior nonsteroidal anti-inflammatory drugs. A large number of research studies aimed at finding selective COX-2 inhibitors were reported earlier [13-16]. Many studies have been carried out using computer simulations to develop rational methods for designing new COX-2 inhibitors [17-27].

Most of those approaches are focused on quantitative structure-activity relationship (QSAR) studies, using different kinds of molecular descriptors for encoding chemical information [17,24,27]. After computing a set of descriptors, multivariate linear or/and nonlinear relationships are established between a reduced subset of variables and the biological activity, leading to a mathematical model [28,29]. The essential characteristic of any QSAR model is, of course, its predictive power. Currently, there are different methods about the description of the molecular structures, including two-dimensional topological descriptors, energetic descriptors, quantum mechanical descriptors, and three dimensional molecular field descriptors [29]. However,

implementing the methods based on 3D structure is in general difficult and time-consuming because of the difficulty in generating optimal 3D conformation of the molecule under consideration. And many current excellent QSAR methods based on two-dimensional properties of the molecule also have a comparable quality to the 3D methods. Again, the true aim of this kind of approach is to quickly build QSAR models, which, if statistically validated to be predictive, will quickly screen *in silico* large databases of small molecules to identify more active molecules.

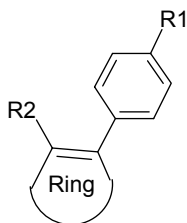
In the present paper, we develop 2D–QSAR models that may be used as screening tools for accurate and rapid quantification and classification of a set of specific COX–2 inhibitors. For that, we have described an interesting application of 2D–autocorrelation vectors for encoding the chemical information derived from a structure diagram. A total of 153 COX–2 inhibitors belonging to diverse chemical families are selected from the literature [22–27] and are used to build two QSAR models using regression and classification techniques. Validation was done with an external test set containing compounds similar to those in the training set [22–27]. This work attempts to test the performance of the 2D autocorrelation method and to evaluate its applicability as a powerful screening method for predicting compounds' activities, given only the 2D molecular structures. Thus, a separate autocorrelation vector was computed for each desired property, and the resulting set was reduced into a smaller number of variables using principal component analysis (PCA), which are used to develop linear discriminant analysis classification models.

2 MATERIALS AND METHODS

2.1 QSAR Dataset

The data set used for this study contains 185 compounds which are selected with purpose of defining a manageable data set characterized by adequate biological and structural diversity [22–27]. Most of them belong to nine chemical families and others are traditional NSAIDs. Molecular structures, numbering of the substituents are enumerated with their generic formula and the original activities in Table 1. As binding data have not been measured in the same experimental conditions, we have given in Table 1, the logarithm of the relative values, noted Ar, which are calculated with respect to the same reference molecule, the celecoxib featured in all the series arising from the different sources [22–27]. The binding affinities measured *in vitro* are in terms of IC₅₀, the drug concentration that inhibits 50% of the COX–2 enzymatic activity. Therefore, for each molecule, $Ar = (IC_{50} \text{ of the molecule}) / (IC_{50} \text{ of the celecoxib})$. A training set containing 153 molecules was used for the generation of QSAR models. A test set constituted by the remaining molecules of the whole database with uniformly distributed biological activities was used to test the predictive ability of the generated models. As celecoxib is the reference molecule, it appears in each of the families given in Table 1 with the original value of IC₅₀. When the references in the same chemical family are different for the IC₅₀ original values it is given in a specific column.

Table 1. Structures of the 185 inhibitors COX-2, original value of IC_{50} , reference and log (Ar), and each molecule of the test set is coded by (T)



1,2-diaryl heterocyclic derivatives[22]

Molecule	Ring	R ₁	R ₂	IC ₅₀ exp/μM	Log(Ar)
1 Celecoxib		SO ₂ NH ₂	4-CH ₃ -C ₆ H ₄	1.000	0.000
2		SO ₂ NH ₂	3,4-di-Cl-C ₆ H ₃	0.398	-0.400
3		SO ₂ CH ₃		9.772	0.990
4		SO ₂ CH ₃	3,5-di-F-C ₆ H ₃	13.490	1.130
5		SO ₂ CH ₃	3,4-di-F-C ₆ H ₃	2.188	0.340
6		SO ₂ CH ₃	3,4-di-F-C ₆ H ₃	17.418	1.241
7		SO ₂ NH ₂		19.099	1.281
8		SO ₂ NH ₂		4.677	0.670
9		SO ₂ CH ₃		33.113	1.520
10		SO ₂ CH ₃	4-OH-C ₆ H ₄	33.113	1.520
11		SO ₂ CH ₃		33.113	1.520

Table 1. (Continued)

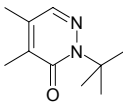
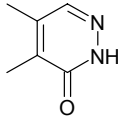
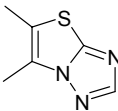
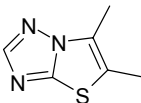
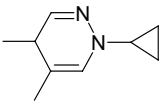
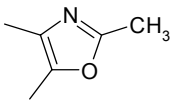

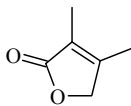
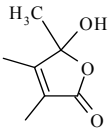
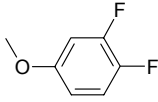
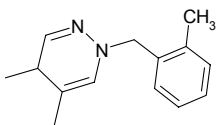
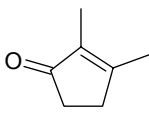
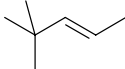
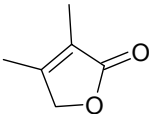
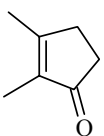
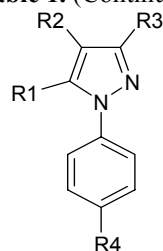
Molecule	Ring	R ₁	R ₂	IC _{50 exp} /μM	Log(Ar)
12		SO ₂ CH ₃	C ₆ H ₅	33.113	1.520
13		SO ₂ CH ₃	C ₆ H ₅	33.113	1.520
14		SO ₂ CH ₃	3,4-di-F-C ₆ H ₃	33.113	1.520
15		SO ₂ CH ₃	C ₆ H ₅	33.113	1.520
16		SO ₂ CH ₃	C ₆ H ₅	33.113	1.520
17		SO ₂ NH ₂		33.113	1.520
18		SO ₂ CH ₃	C ₆ H ₆	0.500	-0.301
19		SO ₂ CH ₃		33.110	1.520
20		SO ₂ CH ₃	C ₆ H ₆	33.110	1.520
21 (T)		SO ₂ CH ₃		5.244	0.720
22 (T)		SO ₂ CH ₃	C ₆ H ₆	100.000	2.000
23 (T)		SO ₂ CH ₃	4-F-C ₆ H ₄	0.08	-1,0969

Table 1. (Continued)



1,5-diaryl-pyrazoles derivatives [23,24,27]

Molecule	R ₁	R ₂	R ₃	R ₄	IC _{50 exp} /μM	Ref.	Log(Ar)
1 Celecoxib		H	CF ₃	SO ₂ NH ₂	0.0398	[23]	
					0.040	[24]	0.000
					0.0398	[27]	
24		H	CF ₃	SO ₂ NH ₂	0.027	[24]	-0.172
25		H	CF ₃	SO ₂ NH ₂	0.066	[24]	0.218
26		H	CF ₃	SO ₂ NH ₂	0.056	[24]	0.148
27		H	CF ₃	SO ₂ NH ₂	0.120	[24]	0.478
28		H	CF ₃	SO ₂ NH ₂	0.058	[24]	0.158
29		H	CF ₃	SO ₂ NH ₂	0.056	[24]	0.148
30		H	CF ₃	SO ₂ NH ₂	0.069	[24]	0.238
31		H	CF ₃	SO ₂ NH ₂	0.288	[24]	0.858
32		H	CF ₃	SO ₂ NH ₂	0.646	[24]	1.208
33		H	CF ₃	SO ₂ NH ₂	0.016	[24]	-0.402
34		H	CF ₃	SO ₂ NH ₂	0.427	[24]	1.028
35		H	CF ₃	SO ₂ NH ₂	0.603	[24]	1.178

Table 1. (Continued)

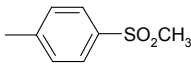
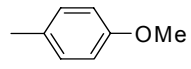
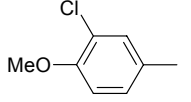
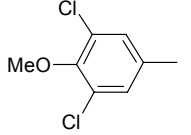
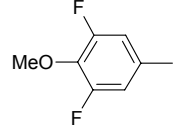
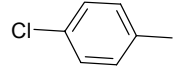
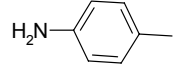
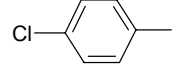
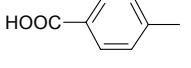
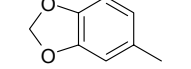
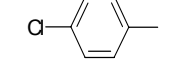
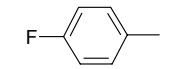
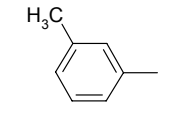
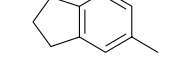
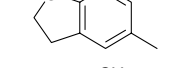
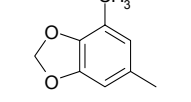
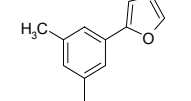
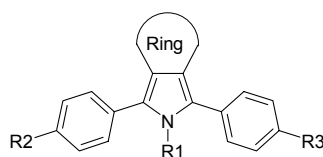
Molecule	R ₁	R ₂	R ₃	R ₄	IC ₅₀ _{exp} /μM	Ref.	Log(Ar)
36		H	CF ₃	SO ₂ NH ₂	100.000	[24]	3.398
37		H	CHF ₂	SO ₂ NH ₂	0.015	[24]	-0.422
38		H	CHF ₂	SO ₂ NH ₂	0.027	[24]	-0.172
39		H	CHF ₂	SO ₂ NH ₂	0.021	[24]	-0.282
40		H	CHF ₂	SO ₂ NH ₂	0.347	[24]	0.938
41		H	CF ₃	SO ₂ NH ₂	0.010	[24]	-0.602
42		H	CF ₃	SO ₂ NH ₂	3.311	[23]	1.918
43		Cl	CF ₃	SO ₂ NH ₂	0.009	[23]	-0.642
44		H	CF ₃	SO ₂ NH ₂	0.005	[23]	-0.872
45		H	CF ₃	SO ₂ NH ₂	11.220	[23]	2.448
46		Me	CF ₃	SO ₂ NH ₂	100.000	[23]	3.398
47		H	COOH	SO ₂ NH ₂	100.000	[23]	3.398
48		H	CF ₃	SO ₂ NH ₂	0.028	[23]	-0.152
49		H	CF ₃	SO ₂ NH ₂	0.110	[23]	0.438
50		H	CF ₃	SO ₂ NH ₂	0.031	[23]	-0.112
51		H	CF ₃	SO ₂ NH ₂	0.008	[23]	-0.702
52		H	CH ₂ F	SO ₂ NH ₂	0.052	[27]	0.118

Table 1. (Continued)

Molecule	R ₁	R ₂	R ₃	R ₄	IC _{50 exp} /μM	Ref.	Log(Ar)
53 (T)		H	CF ₃	H	0.066	[27]	0.220
54 (T)		Cl	COOH	SO ₂ NH ₂	71.150	[27]	3.250
55 (T)		OH	CF ₃	SO ₂ NH ₂	3.548	[27]	1.948
56 (T)		H	CH ₂ F	SO ₂ NH ₂	100.000	[27]	3.398



1,3-diaryl heterocyclic derivatives[22]

molecule	Ring	R ₁	R ₂	R ₃	IC _{50 exp} /μM	Log(Ar)
1 Celecoxib					1.000	0.000
57		H	H	H	0.015	-1.824
58		H	H	H	0.033	-1.481
59		NH ₂	H	H	0.100	-1.000
60		H	CH ₃ S	CH ₃ S	0.501	-0.300
61		H	F	F	0.017	-1.770
62		H	H	H	0.021	-1.678
63 (T)		H	4-F-3-NHAc	4-F-3-NHAc	0.005	-2.301
64		H	F	Imidzol-1-yl	0.042	-1.377
65		H	H	H	0.014	-1.854

Table 1. (Continued)

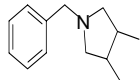
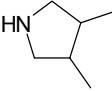












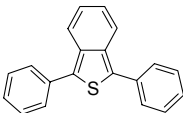
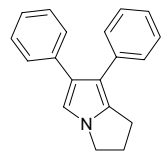
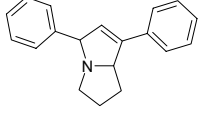
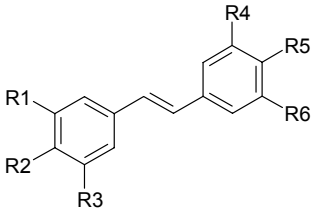
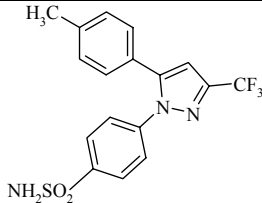
molecule	Ring	R ₁	R ₂	R ₃	IC ₅₀ _{exp} /μM	Log(Ar)
66		H	H	H	0.100	-1.000
67		H	H	H	0.100	-1.000
68 (T)		H	H	H	0.003	-2.523
69		H	F	H	0.004	-2.398
70		H	F	F	0.708	-0.150
71		H	CH ₃ SO ₂	CH ₃ SO ₂	10.000	1.000
72		H	H	CH ₃ SO ₂	0.002	-2.699
73		H	H	H	0.035	-1.456
74		H	4-F-3-NHAc	H	0.100	-1.000
75		H	4-F-3-NHCH ₃	4-F-3-NHAc	0.100	-1.00
76		H	F	F	0.120	-0.921
77		H	4-F-3-NHCH ₃	4-F-3-NHCH ₃	0.100	-1.000
78 (T)		CH ₃	H	H	1.000	0.000
79 (T)		H	F	F	0.080	-1.097
80					0.011	-1.959
81					0.029	-1.538
82 (T)					0.050	-1.301

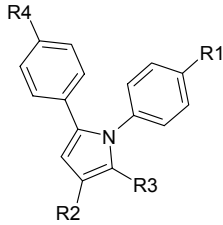
Table 1. (Continued)

Resveratrole derivatives [25]



molecule	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	IC ₅₀ exp/μM	Log(Ar)
1 Celecoxib							0.035	0.000
83	OCH ₃	H	OCH ₃	H	OCH ₃	H	1.676	1.680
84	OCH ₃	OCH ₃	OCH ₃	H	OCH ₃	H	7.837	2.350
85	OCH ₃	H	OCH ₃	OCH ₃	H	OCH ₃	1.583	1.655
86	OCH ₃	H	OCH ₃	OCH ₃	OCH ₃	H	0.800	1.359
87	OCH ₃	OCH ₃	OCH ₃	OCH ₃	H	OCH ₃	0.517	1.169
88	OH	H	OH	H	OH	H	1.001	1.456
89	OH	OH	OH	H	OH	H	0.046	0.115
90	OH	H	OH	OH	OH	H	0.0113	-0.489
91	OH	OH	OH	OH	H	OH	0.00138	-1.402
92	OH	OH	OH	OH	OH	OH	0.00104	-1.525
93	OH	H	OCH ₃	H	OH	H	2.221	1.803
94	OCH ₃	H	OCH ₃	H	OH	H	1.196	1.534
95	OH	H	OH	OH	H	OH	0.00171	-1.301

1,5-diarylpyrrole-3-acetic derivatives [26,27]



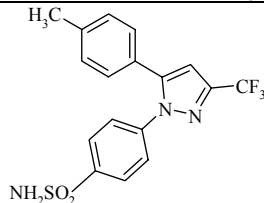
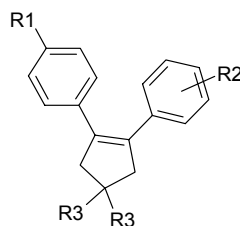
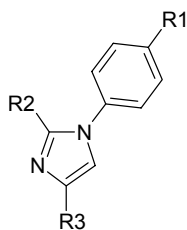
molecule	R ₁	R ₂	R ₃	R ₄	IC ₅₀ exp/μM	Ref.	Log(Ar)
1 Celecoxib					0.079	[26]	0.000
					0.0398	[27]	
96	H	COCOOC ₂ H ₅	CH ₃	SO ₂ CH ₃	4.300	[26]	1.736
97	CF ₃	COCOOC ₂ H ₅	CH ₃	SO ₂ CH ₃	9.900	[26]	2.098
98	CF ₃	CH ₂ COOC ₂ H ₅	CH ₃	SO ₂ CH ₃	0.060	[26]	-0.119
99	CH ₃	CH ₂ COOC ₂ H ₅	CH ₃	SO ₂ CH ₃	0.480	[26]	0.784
100	H	CH ₂ COOH	CH ₃	SO ₂ CH ₃	1.000	[26]	1.102
101	CH ₃	CH ₂ COOH	CH ₃	SO ₂ CH ₃	0.430	[26]	0.736
102	CH ₃	CH ₂ COOH	CH ₃	SO ₂ CH ₃	0.110	[26]	0.144
103	SO ₂ NH ₂	H	H	4-F	0.014	[27]	-0.452
104	4-F	H	CH ₃	4-SO ₂ CH ₃	0.060	[27]	0.179
105	4-F	CH ₂ OCOCH ₃	CH ₃	4-SO ₂ CH ₃	0.470	[27]	1.070
106 (T)	4-F	CH ₂ OH	CH ₃	4-SO ₂ CH ₃	3.910	[27]	1.990
107 (T)	4-F	CH ₂ CF ₃	CH ₃	4-SO ₂ CH ₃	0.142	[27]	0.550
108 (T)	H	H	CH ₃	4-SO ₂ CH ₃	0.061	[27]	0.180
109 (T)	4-F	CHO	CH ₃	4-SO ₂ CH ₃	3.252	[27]	1.910

Table 1. (Continued)



Cyclopentene derivatives [27]					
molecule	R ₁	R ₂	R ₃	IC ₅₀ exp/μM	Log(Ar)
1 Celecoxib				0.0398	0.000
110	SO ₂ CH ₃	3,4-diF	H	0.051	0.108
111	SO ₂ CH ₃	4-F-3-Cl	H	0.030	-0.122
112	SO ₂ CH ₃	3,4-diCl	H	0.010	-0.602
113	SO ₂ CH ₃	4-OCH ₃ -3-F	H	0.120	0.478
114	SO ₂ CH ₃	4-OCH ₃ -3,5-diCl	H	0.017	-0.372
115	SO ₂ CH ₃	3,4-OCH ₂ CH ₂ O	H	0.021	-0.282
116	SO ₂ CH ₃	4-N(CH ₃) ₂ -3-Cl	H	0.005	-0.902
117	SO ₂ NH ₂	4-F	H	0.007	-0.752
118	SO ₂ NH ₂	3,4-diF	H	0.018	-0.342
119	SO ₂ NH ₂	4-F-3-Cl	H	0.010	-0.602
120	SO ₂ NH ₂	4-Cl	H	0.003	-1.122
121	SO ₂ NH ₂	3,4-diCl	H	0.002	-1.302
122	SO ₂ NH ₂	3-F-4-OCH ₃	H	0.016	-0.402
123	SO ₂ NH ₂	3-Cl-4-OCH ₃	H	0.009	-0.652
124	SO ₂ NH ₂	3,5-diCl-4-OCH ₃	H	0.009	-0.652
125	SO ₂ NH ₂	4-CF ₃	H	0.151	0.578
126	SO ₂ NH ₂	3-F-4-CF ₃	H	0.170	0.628
127	SO ₂ CH ₃	4-OCH ₃	H	0.005	-0.902
128	SO ₂ CH ₃	4-Cl	H	0.003	-1.122
129	SO ₂ CH ₃	4-CH ₃	H	0.003	-1.122
130	SO ₂ CH ₃	4-CF ₃	H	0.052	0.118
131	SO ₂ CH ₃	4-SCH ₃	H	0.219	0.738
132	SO ₂ CH ₃	2-CH ₃ -4-F	H	0.076	0.278
133	SO ₂ CH ₃	4-F	CH ₃	0.015	-0.422
134	SO ₂ CH ₃	4-Cl	CH ₃	0.007	-0.752
135	SO ₂ CH ₃	4-F	CF ₃	0.068	0.228
136	SO ₂ CH ₃	4-F	CH ₂ F	0.051	0.108
137	SO ₂ NH ₂	4-OCH ₃	H	0.002	-1.301
138	SO ₂ NH ₂	4-N(CH ₃) ₂ -3-Cl	H	0.002	-1.301
139	SO ₂ CH ₃	4-CH ₂ OCH ₃	H	6.610	2.220
140 (T)	SO ₂ CH ₃	4-OCH ₃ -3-Cl	H	0.142	0.550
141 (T)	SO ₂ NH ₂	3,5-diCl-4-OCH ₃	H	0.009	-0.650
142 (T)	SO ₂ CH ₃	4-F	H	0.026	-0.190
143 (T)	SO ₂ CH ₃	4-CN	H	3.252	1.910

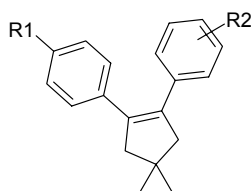
Table 1. (Continued)



Imidazole derivatives [27]					
molecule	R ₁	R ₂	R ₃	IC ₅₀ _{exp} /μM	Log(Ar)
1 Celecoxib				0.0398	0.000
144	SO ₂ CH ₃		CF ₃	0.158	0.598
145	SO ₂ NH ₂		CF ₃	0.010	-0.602
146	SO ₂ NH ₂		CF ₃	0.010	-0.602
147	SO ₂ NH ₂		CF ₃	0.040	-0.002
148	SO ₂ CH ₃		CH ₂ CN	1.556	1.590
149	SO ₂ CH ₃		CF ₃	0.121	0.480
150	SO ₂ NH ₂		CF ₃	0.040	0.000
151 (T)	SO ₂ CH ₃		CF ₃	1.707	1.630
152 (T)	SO ₂ CH ₃		CF ₃	9.598	2.380
153 (T)	SO ₂ CH ₃		CF ₃	1.829	1.660
154	SO ₂ CH ₃		CF ₃	1.829	1.660

Table 1. (Continued)

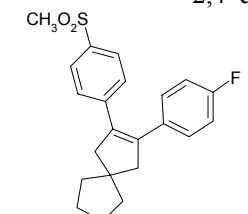
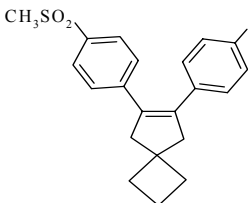
molecule	R ₁	R ₂	R ₃	IC _{50 exp} /μM	Log(Ar)
155 (T)	SO ₂ CH ₃		CF ₃	1.208	1.480
156	SO ₂ CH ₃		CF ₃	38.210	2.980
157 (T)	SO ₂ CH ₃		CF ₃	0.960	1.380
158 (T)	SO ₂ NH ₂		CF ₃	0.341	0.930
159 (T)	SO ₂ CH ₃		CF ₃	3.252	1.910
160 (T)	SO ₂ CH ₃		CF ₃	0.917	1.360
161 (T)	SO ₂ CH ₃		CF ₃	0.333	0.920
162 (T)	SO ₂ CH ₃		CF ₃	1.052	1.420
163 (T)	SO ₂ CH ₃		CF ₃	0.142	0.550



Spiroheptene derivatives [27]

molecule	R ₁	R ₂	IC _{50 exp} /μM	Log(Ar)
1 Celecoxib			0.0398	0.000
164	SO ₂ NH ₂	3-Br-4-OCH ₃	0.002	-1.302
165	SO ₂ CH ₃	3,5-diCl-4-OCH ₃	0.006	-0.822
166	SO ₂ NH ₂	3,5-diCl-4-OCH ₃	0.004	-1.002
167	SO ₂ CH ₃	3-Cl-4-F	0.007	-0.752
168	SO ₂ NH ₂	3-Cl-4-F	0.002	-1.422
169	SO ₂ CH ₃	3-Cl-4-OCH ₃	0.017	-0.372

Table 1. (Continued)

molecule	R ₁	R ₂	IC ₅₀ exp/μM	Log(A _r)
170	SO ₂ NH ₂	3-Cl-4-OCH ₃	0.002	-1.302
171	SO ₂ CH ₃	2,4-diCl	0.033	-0.082
172			0.062	0.188
173 (T)			0.004	-1.000

Structures of traditional NSAIDs with their original IC₅₀ and log(A_r) studied [22]

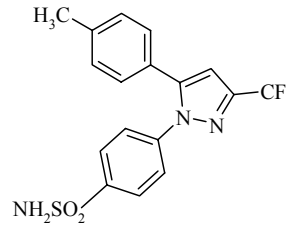
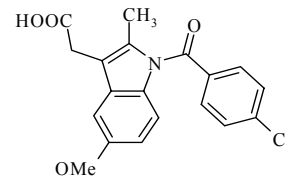
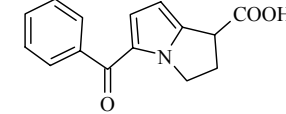
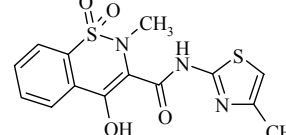
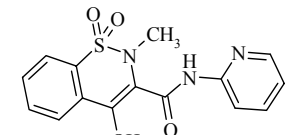
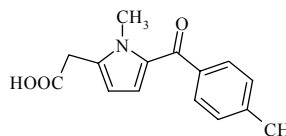
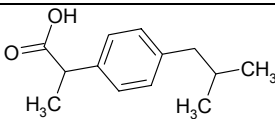
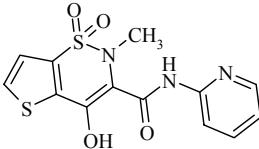
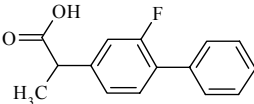
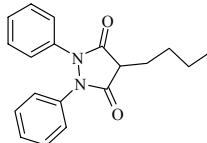
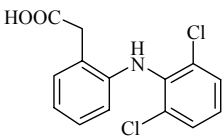
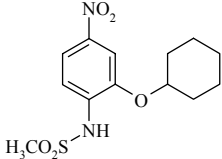
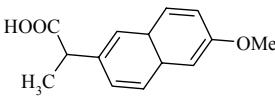
molecule	Structure	IC ₅₀ exp/μM	Log(A _r)
1 Celecoxib		1.000	0.000
174 Indomethacin		0.46	-0.340
175 Ketorolac		0.86	-0.060
176 Meloxicam		0.75	-0.120
177 Piroxicam		8.99	0.950
178 Tolmetin		7.09	0.850

Table 1. (Continued)

molecule	Structure	IC ₅₀ _{exp} /μM	Log(Ar)
179 Ibuprofen		30.20	1.480
180 Tenoxicam		14.22	1.149
181 Flurbiprofen		6.46	0.810
182 Phenylbutazone		30.2	1.480
183 Diclofenac		0.05	-1.301
184 NS-398		0.47	-0.328
185 Naproxen		73.74	1.868

2.2 Computational Procedures

2.2.1 2D Molecular structure building and calculation of descriptors

The 2D structures of the 185 COX-2 inhibitors were drawn with standard geometric parameters using the molecular builder of ISIS/Draw as implemented in the ISIS 2.4 package. All structures were stored after concatenation in a molfile (MDL format) with the relative activities.

2D autocorrelation vector is presented as an intrinsic descriptor of the distribution of an atomic property along the molecular graph. The first application of these vectors as molecular descriptors has been published in 1980, by Moreau and Broto [30,31], who applied the classical mathematical notion of an autocorrelation function to the topology of molecular structures. For these vectors, H-depleted molecular structure is represented as a graph G and physicochemical properties of atoms as real values assigned to the vertices of G . These descriptors can be obtained by summing up the products of certain properties of two atoms, located at given topological distances. Given that each atom, I , has an associated property $f(I)$, an autocorrelation vector, V , is defined such that:

$$V = (V_1, V_2, \dots, V_j, \dots, V_{\max})$$

where V_j contains the sum of products $f[K] \times f[L]$ for all pairs of atoms, K and L, separated by exactly by j bonds, $0 \leq j \leq \max$, and \max is the maximum possible interatomic separation within that structure.

Autocorrelation vectors represent the degree of similarity between molecules and offer several advantages, in that they are compact, independent of the original atom numbering, and their length is independent of the size of the molecule. The size of an autocorrelation vector is determined by the largest path between two atoms, so that different molecules usually have vectors having different numbers of components, making their comparison and the statistical analysis more difficult. To overcome this difficulty, we used a slight variation of autocorrelation: instead of considering only the shortest path between two atoms, all sorts of path were considered: they are easily obtained by the successive powers of the connectivity matrix M : in M^x the number of path of length x bonds between atoms K and L is given by the value of the term $M^x(K,L)$. Clearly, in this variation, V_j contains the sum of products $f[K] \times f[L]$ for all pairs of atoms, K and L, separated by J bonds, in any sort of path (with return, cyclic, etc...).

Then, using autocorrelation2 program [32], we calculated 2D autocorrelation vectors. The eight atomic properties included in – or calculable by – the program and were used as follows: Pauling electronegativities, van der Waals volumes, indicator for saturated/unsaturated, logP contribution, polar surface area, indicator for hydrogen–bond donor, indicator for hydrogen–bond acceptor and partial charge.

Then, a data matrix is generated with the autocorrelation vectors calculated for each compound. A total of 120 components were computed for each compound, derived from 8 atomic properties and limited to path of length 0 to 14 bonds. The components of these vectors are the descriptors of the molecules and those with constant values were discarded. Afterwards, dimensionality reduction methods are performed for selecting the most relevant vector components for building quantitative and qualitative models.

In spite of the obvious limited information from 2D representation of molecules, 2D autocorrelation descriptors have already found successful applications in the performance of Artificial Neural Networks based QSAR modeling [33,34].

2.2.2 Statistical methods

2.2.2.1 Multiple Linear Regression (MLR)

On the whole autocorrelation vectors calculated there are some strong correlations between them. So, it is impossible to directly use MLR to develop a QSAR model, and it is essential to employ the multivariable statistical methods based on factor analysis such as principal component

analysis (PCA) to extract non-redundant information.

Initially, the data analysis was performed by PCA using the ACPLPK2 [35] package. Technically speaking, PCA is an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA can be used for dimensionality reduction in a data set while retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the most important aspects of the data. But this is not necessarily the case, depending on the application. PCA is a useful exploratory tool, which maps samples through scores and individual variables by the loadings in a new vector space defined by the principal components [36].

PCA calculates the covariance matrix, and then computes the eigenvalues and eigenvectors of the covariance matrix. The eigenvectors that correspond to the highest valued eigenvalues are called the principal components (PCs) of the covariance matrix. This modification ensures the comparable importance of the variables and avoids the risk that a small but systematic variation in one variable can be masked by a large variation in another. In Figure 1, we can see that the training and test sets are widely and homogeneously spread in the plane defined by the two main PCA axes calculated from the autocorrelation vectors describing the training set.

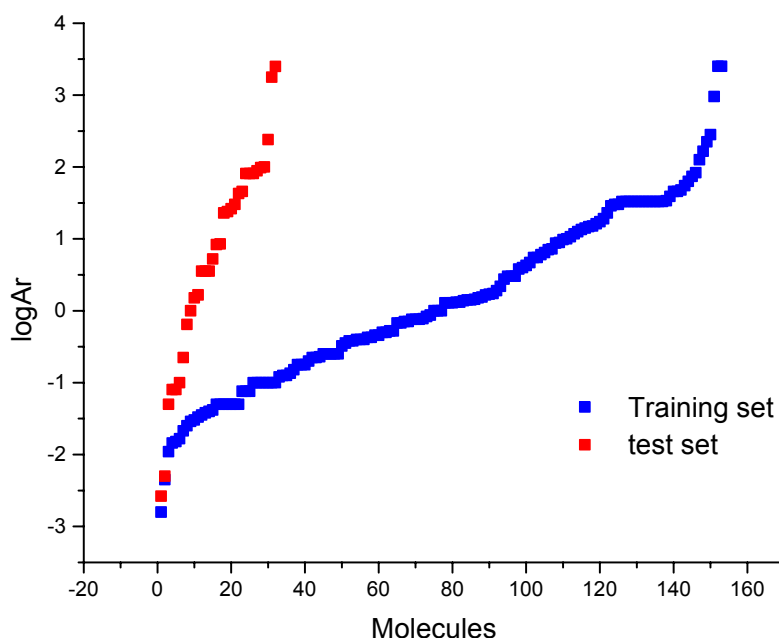


Figure 1. COX-2 activity distributions of the 153 molecule training set and the 32 molecule test set.

After the PCs have been computed, the quantitative QSAR models are built from the training set, using the REGMUL [37] program which allows to find multivariate linear relationships between PCs and the biological activity (related here by means of relative value). This program obeys to

statistical criteria chosen by the user, which are: number of variables in the linear relationship, number of PCs among which they are chosen (the n first...), the iteration count to be used by the Monte Carlo method, included in the program, which chooses by chance the variables of the regression. The procedure is repeated by changing criteria's values and until the best statistical model is obtained. Afterwards, we then looked at the predictive powers of the obtained model relative to 32 inhibitors of the test set.

2.2.2.2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is another commonly used technique for data classification and dimensionality reduction. LDA easily handles the case where the within-class frequencies are unequal and their performances have been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. The prime difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification. In PCA, the shape and location of the original data sets changes when transformed to a different space whereas LDA doesn't change the location but only tries to provide more class separability and draw a decision region between the given classes. This method also helps to better understand the distribution of the feature data.

The qualitative model is built from the same training set, each ligand being described by the PCs and one COX-2 inhibition class, either the active class or inactive class if the ligand's COX-2 Ar is, respectively, lower than 10.680 or greater than. LDA was performed, using the ANADIS2 [37] package, on both classes of training set. The discriminating ability is evaluated by the percentage of correct classifications into each class.

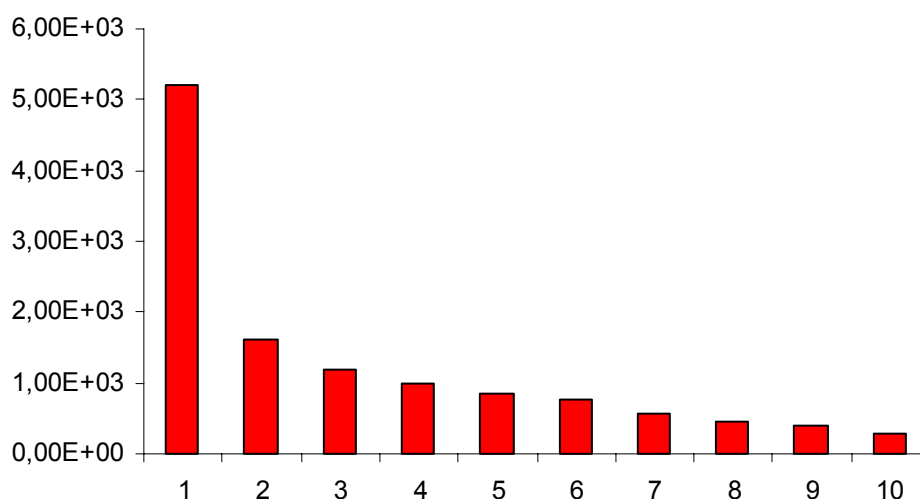


Figure 2. Diagram of the first ten eigenvalues of the ACP.

3 RESULTS AND DISCUSSION

Once the quantitative and qualitative models were built, each molecule of the training and test sets has its COX–2 activity predicted. These predictions are the basis of the results presented in this paper. The application of the PCA on the autocorrelation vectors generates principal components (PCs), ordered by decreasing eigenvalues of the covariance data matrix. In Figure 2, the first component represents the largest variance in the data set, the second component the largest of the remaining variance, and so on. The major patterns within the original data are thus captured by a small number of components. In another terms, the PCA factors n°1 and n°2 encode almost all the information. Figure 3 shows the distribution of molecules in the first plane, each point corresponds to a molecule of the whole set.

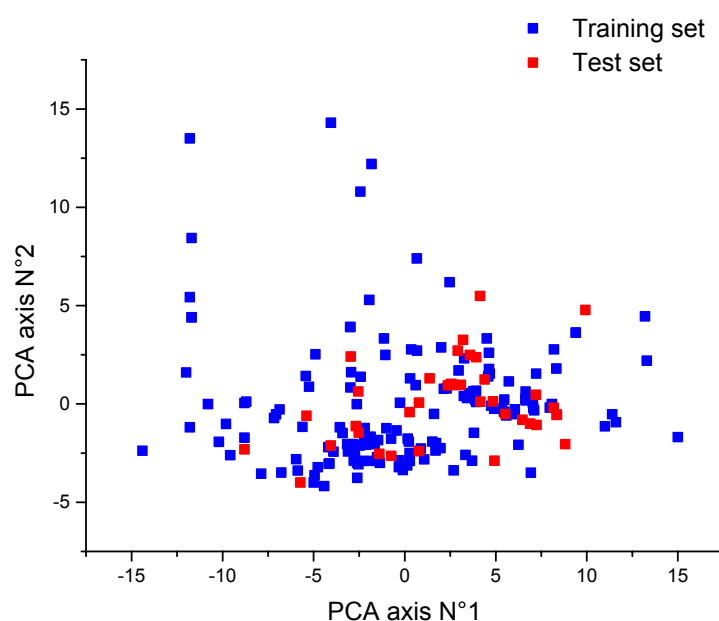


Figure 3. Plane created by the two first factors from PCA on the 120 descriptors of the 153 COX–2 inhibitors. The molecules of the training set are shown as blue squares, and the molecules of test set are shown as red squares. The PCA factors n°.1 and n°. 2 account for, respectively, 42% and 13% of the whole molecular information.

3.1 Multiple Linear Regression

The MLR is employed to develop the best QSAR model for the COX–2 inhibitors belonging to chemical families considered in this work. This method was used to generate a ten–variable linear model for the anti COX–2 activity. The validity of the model was proven by the multiple correlation coefficient ($R = 0.890$) and the F test value. In addition, this model is very satisfying since the entire set is characterized by a high degree of structural variety. Plots in Figure 4 depict the fitting of the training and test sets. The predictive power of the proposed model was discerned by successfully testing the 32 compounds constituting the test set. The results are discussed below.

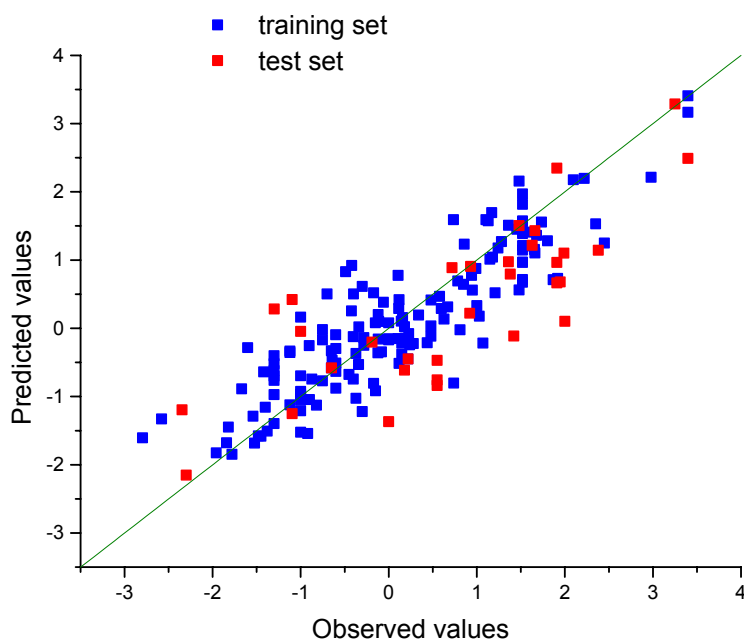


Figure 4. Plot of $\log(A_R)$ values for COX-2 enzyme predicted against those observed for the molecules of the training set and test set.

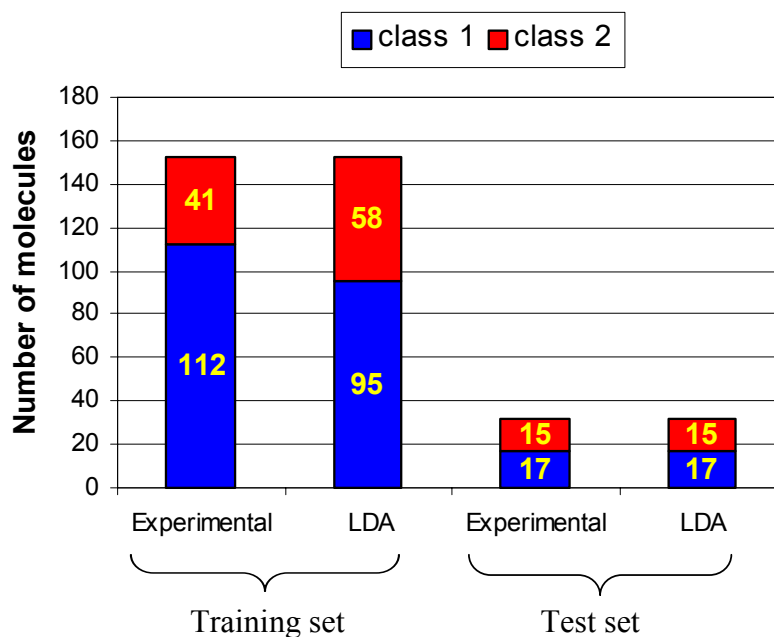


Figure 5. LDA results for the training and test set. Class 1 for the active class and class 2 for the inactive.

3.2 Linear Discriminant Analysis

In this part of present work, an SAR model based on 2D autocorrelation vectors, for encoding the chemical information derived from a plane structure diagram, was developed to predict new molecules with COX-2 inhibition class. The results obtained by this method are represented in the Figure 5. The most important criterion for the acceptance or not of a discriminant model, such as the model developed here, is based on the statistics for the test set. It has been known that a good

QSAR model should possess not only a good calibrated statistics for the internal molecules but also a high predictive ability for the external molecules. In Figures 6 and 7, we can see in details how well both COX-2 models have been trained by looking at the distribution of the predictions for the training set and the test set, respectively. A molecule will be predicted as active if a quantitative model predicts its COX-2 activity inhibition, related here by Ar , lower than **10.680**, or if a qualitative model classifies this molecule into the active category. A COX-2 inhibitor can then be classified into four types depending on its predicted and real COX-2 activity: True Positive (TP), False Positive (FP) when an inactive product is predicted positive, True Negative (TN), or False Negative (FN) when an active compound is predicted inactive.

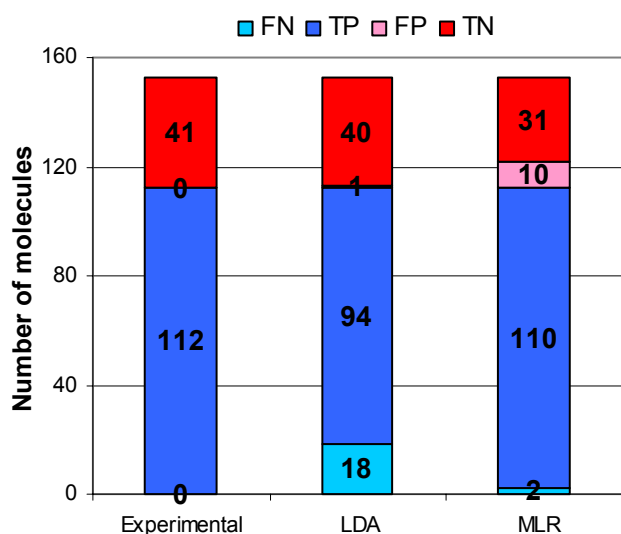


Figure 6. Classification of the predictions made by the MLR COX-2 model and the LDA COX-2 model on the 153 compound training set.

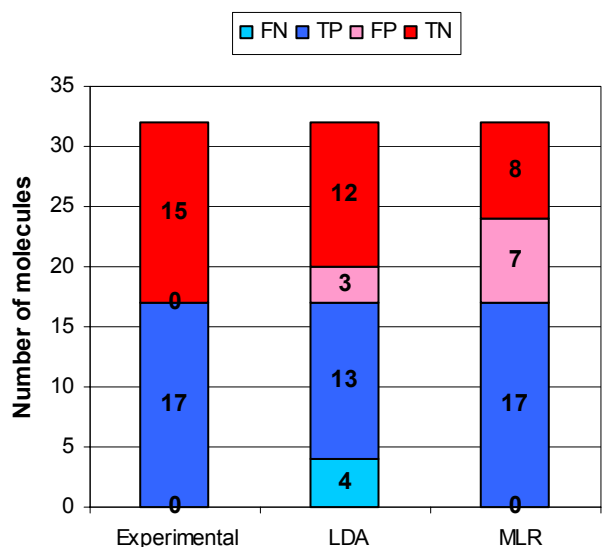


Figure 7. Classification of the predictions made by the quantitative COX-2 model and the qualitative COX-2 model on the 32 compound test set.

Figures 6 and 7 show that the LDA method has higher predictive ability for the molecules of the training and test set than MLR method for the inactive class, and a smaller predictive ability than MLR method for the active class. Thus, the LDA results and the MLR results are complementary. With regard to the training set, the molecules in the inactive class are classified according to a percentage of 97.5% of correct classification and the molecules in the active class are predicted according to a percentage of 98.2% of correct prediction. As our aim is to use LDA for not keeping inactive molecules and MLR to detect active molecules these results validate the models for use in the ligand-based virtual screening taking into consideration that 85% is considered as an acceptable threshold limit for this kind of analysis [38].

The molecules of the test set are distributed in the two classes: the molecules of the active class are classified according to a percentage of 77% of correct classification while those which must belong to the inactive class to 80% are well classified. Specifically, the model correctly classifies 13 out of 17 anti-inflammatory compounds and 12 out of 15 inactive chemicals in the test set.

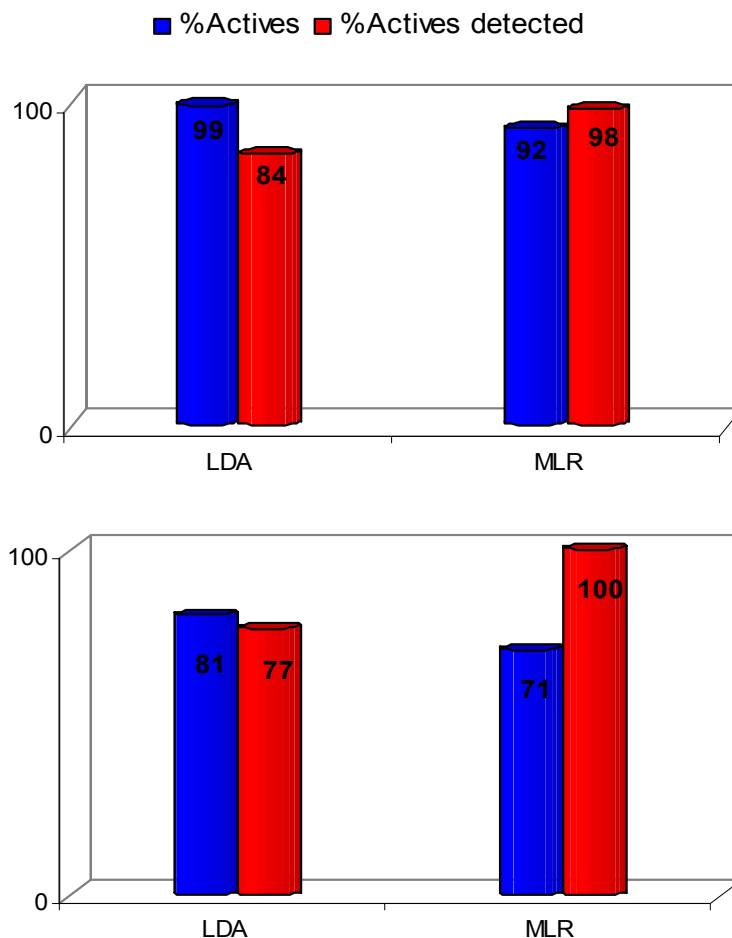


Figure 8. Composition of the subset selected by each model from the 153 ligand training set (top) and from the 32 ligand test set (bottom).

In order to compare in details the predictive powers of obtained classification and regression models, we report in Figure 8, two parameters: % Actives and % Actives detected, which can be calculated from the MLR model as well as from the LDA model. % Actives is the percentage of real active inhibitors in the subset predicted as active, and it is defined as $TP/(TP+FP)$. % Actives detected is the percentage of real active inhibitors detected among all the real active inhibitors, and it is defined as $TP/(TP+FN)$. Of course, a perfect model would detect all the real active inhibitors and without mistakes (both parameters are equal to 100%).

Figure 8 shows the power for both COX-2 models to enrich the selected subset with active inhibitors, while detecting as many as possible of all the active COX-2 inhibitors. The LDA and MLR models are good since they predict as active a subset that contains 77% and 100% of active inhibitors and that contains, respectively, 84% and 98.2% of all the active inhibitors from the training set. Interestingly, the trends are similar to what was observed when looking at the predictive performances on the test set. As the molecules of test set were not used for the training of the models, we simulated here the practical application of the regression and the classification

models during a virtual screening when they must detect a maximum of active COX–2 inhibitors while being wrong as rarely as possible.

It is important to note that Estrada and co–workers have shown that new lead drug can be designed and/or selected even if its mechanism of action is completely unknown, by using algorithms based on the structural characterization of a structurally diverse database with molecular descriptors and some pattern recognition technologies such as LDA [39–40]. Thus, the use of LDA as a first filter within the framework of the virtual screening of selective inhibitors of COX–2 leads to a very significant improvement for the inactive set. As the goal is to minimize the risk of selecting false active candidates, this is the best strategy. Anyway, this LDA model should work as a very good filter for the selection of new COX–2 inhibitors. Then, it is interesting to use MLR to seek the most active molecules.

4 CONCLUSIONS

Biological phenomena are complex by nature. In this work, the inhibitory activity against COX–2 enzyme of a set of compounds was successfully modelled using MLR and LDA. 2D autocorrelation descriptors were used for encoding structural information of the studied compounds. The most predictive model was generated with MLR, although LDA produced satisfactory statistical results for the inactive set. The robustness and predictive ability of the models were evaluated by selecting a wide range of biological activities with chemical classes similar to the training set. The above results show that the models developed in this paper have good calibrated statistics and high predictive ability. The external predicting series (test set) revealed that the QSAR and the SAR models had a good predictability. Thus, this study demonstrates that 2D autocorrelation is a useful methodology for identifying new compound with anti–inflammatory activity and it represents a novel and rather promising way to bioinformatics research.

Finally, we conclude that our approach, based on 2D–structure so rapidly computed but low–information–content descriptors, provides an excellent tool for the prediction of COX–2 inhibition activity. It uses the complementarities of LDA and MLR methods to enrich a set of molecules with putative active ones and may be a fruitful preliminary step for a docking approach. Thus, the data generated should be useful for predictive purposes, as an *in silico* filter, to high–throughput virtual screening of the large structural database for novel, potent and selective COX–2 inhibitor.

Acknowledgment

S. Taïri–Kellou gratefully thanks Dr Bernard Maignet and his group (CNRS, Nancy) for their helpful contributions.

5 REFERENCES

- [1] J. R. Vane, Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs, *Nat. New Biol.* **1971**, *231*, 232–235.
- [2] J.R. Vane and R. M. Botting, Mechanism of action of anti-inflammatory drugs, *Scand. J. Rheumatol. Suppl.* **1996**, *102*, 9–21.
- [3] L.J. Marnett and A. S. Kalgutkar, Cyclooxygenase 2 inhibitors: Discovery, Selectivity and the future, *Trends Pharm. Sci.* **1999**, *20*, 465–469.
- [4] M. C. Allison, A. G. Howatson, C. J. Torrance, F. D. Lee and R. I. Russel, Gastrointestinal Damage Associated with the Use of Nonsteroidal Anti-inflammatory Drugs, *New Engl. J. Med.* **1992**, *327*, 749–754.
- [5] J. J. Talley, Selective Inhibitors of cyclooxygenase-2 (COX-2), *Prog. Med. Chem.* **1999**, *36*, 201–234.
- [6] E. J. Topol, Failing the public health – Rofecoxib, Merck and the FDA, *New Engl. J. Med.* **2004**, *351*, 1707–1709.
- [7] P. Prasit, Z. Wang, C. Brideau, C. C. Chan, S. Charleson, W. Cromlish, D. Ethier, J. F. Evans, A. W. Ford-Hutchinson, J. Y. Gauthier, R. Gordon, J. Guay, M. Gresser, S. Kargman, B. Kennedy, Y. Leblanc, S. Léger, J. Mancini, G. P. O'Neill, M. Ouellet, M. D. Percival, H. Perrier, D. Riendeau, I. Rodger, P. Tagari, M. Yhérien, P. Vickers, E. Wong, L.-J. Xu L.-J, R. N. Young and R. Zambouni, The discovery of rofecoxib, [MK 966, Vioxx, 4-(4'-methylsulfonylphenyl)-3-phenyl-2(5H)-furanone], an orally active cyclooxygenase-2-inhibitor, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1773–1778.
- [8] G. A. FitzGerald, Coxibs and Cardiovascular Disease, *N. Engl. J. Med.* **2004**, *351*, 1709–1711.
- [9] D. H. Solomon, S. Schneeweiss, R. J. Glynn, Y. Kiyota, R. Levin, H. Mogun and J. Avron, Relationship between Selective Cyclooxygenase-2 Inhibitors and Acute Myocardial Infarction in Older Adults, *Circulation* **2004**, *109*, 2068–2073.
- [10] N. M. Davies and F. Jamali, COX-2 Selective Inhibitors Cardiac Toxicity: Getting to the Heart of the Matter, *J. Pharm. Pharm. Sci.* **2004**, *7*, 332–336.
- [11] A. Finckh and M. D. Aronson, Cardiovascular Risks of Cyclooxygenase-2Inhibitors: Where We Stand Now, *Ann. Intern. Med.* **2005**, *142*, 212–214.
- [12] J. M. Dogne, C. T. Supuran and D. Pratico, Adverse Cardiovascular Effects of the Coxibs, *J. Med. Chem.* **2005**, *48*, 2251–2257.
- [13] R. G. Kurumbail, A. M. Stevens, J. K. Gierse, J. J. McDonald, R. A. Slegeman, J. Y. Pak, D. I. Gildehaus, J. M. Miyashiro, T. D. Penning, K. Siebert, P. C. Isakson and W. C. Stallings, Structural basis of selective inhibition of COX-2 by anti-inflammatory agents, *Nature* **1996**, *384*, 644–648.
- [14] Y. Leblanc, W. C. Black, C. C. Chan, S. Charleson, D. Delorme, D. Denis, J. Y. Gauthier, E. L. Grimm, R. Gardon, D. Guay, P. Hamel, S. Kargman, C. K. Lau, J. I. Mancini, M. Ouellet, D. Percival, P. Roy, K. Skorey, P. Tagari, E. Wong, L. Xu and P. Prasit, Synthesis and biological evaluation of both enantiomers of L-761000 as inhibitors of cyclooxygenase-1 and 2, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 731–736.
- [15] K. M. Woods, R. W. McCroskey and M. R. Michaelides, Heterocyclic compounds as COX-2 inhibitors, *World Patent WO989330*, 4 March **1997**.
- [16] T. Klein, R. M. Nusing, J. Pfeilschifetr and V. Ulrich, Selective inhibition of COX-2, *Biochem. Pharmacol.* **1994**, *48*, 1605–1610.
- [17] P. Chavatte, S. Yous, C. Maro, N. Baurin and D. Lesieur, Three dimensional quantitative structure activity relationships of cyclooxygenase-2: A comparative molecular field analysis, *J. Med. Chem.* **2001**, *44*, 3223–3230.
- [18] J. R. Desiraju, B. Gopalkrishnan, R. K. R. Jetti, A. Nagaraju, D. Ravindra, J. R. P. Sarma and M. E. Sobhia, Computer aided drug design of selective COX-2 inhibitors: CoMFA, CoMSIA and docking studies of some 1,2 diarylimidazole derivatives, *J. Med. Chem.* **2002**, *45*, 4847–4855.
- [19] H. Liu, X. Q. Huang, S. J. Shen, X. M. Luo, M. H. Li, B. Xiong, G. Chen, J. K. Shen, Y. M. Yang and K. X. Chen, Inhibitory mode of 1,5 diarylpyrazole derivatives against cyclooxygenase-2 and cyclooxygenase-1, *J. Med. Chem.* **2002**, *45*, 4816–4827.
- [20] R. Solvia, C. Almansa, S. G. Kalko, S. J. Luque and M. Brozco, Theoretical studies on the inhibition mechanisms of COX-2: Is there a unique recognition site? *J. Med. Chem.* **2003**, *46*, 1372–1377.
- [21] R. Garg, A. Kurup, S. B. Mekapati and C. Hansch, Cyclooxygenase inhibition: A comparative QSAR study, *Chem. Rev.* **2003**, *102*, 703–732.
- [22] A. K. Chakraborti, and R. Thilagavathi, Computer-Aided Design of Selective COX-2 Inhibitors: Molecular Docking of Structurally Diverse Cyclooxygenase-2 Inhibitors using FlexX, *Internet Electronic Journal of Molecular Design* **2004**, *3*, 704–719.
- [23] R. Thilagavathi and A. K. Chakraborti, Importance of Alignment in Developing 3-D QSAR Models of 1,5-Diaryl Pyrazoles for Prediction of COX-2 Inhibitory Activity, *Internet Electronic Journal of Molecular Design* **2003**, *2*, 000–000.
- [24] G. R. Desiraju, B. Gopalakrishnan, R. K. R. Jetti, D. Raveendra, Jagarlapudi A. R. P. Sarma, and H. S.

- Subramanya, Three–Dimensional Quantitative Structural Activity Relationship (3D–QSAR) Studies of Some 1,5–Diarylpyrazoles: Analogue Based Design of Selective Cyclooxygenase–2 Inhibitors, *Molecules*, **2000**, *5*, 945–955.
- [25] M. Murias, N. Handler, T. Erker, K. Pleban, G. Ecker, P. Saiko, T. Szekeres and W. Jäger, Resveratrol analogues as selective cyclooxygenase–2 inhibitors: synthesis and structure–activity relationship, *Bioorg. Med. Chem.* **2004**, *12*, 5571–5578.
- [26] M. Biava, G. Cesare Porretta, A. Cappelli, S. Vomero, F. Manetti, M. Botta, L. Sautebin, A. Rossi, F. Makovec and M. Anzini, 1,5–Diarylpyrrole–3–acetic Acids and Esters as Novel Classes of Potent and Highly Selective Cyclooxygenase–2 Inhibitors, *J. Med. Chem.* **2005**, *48*, 3428–3432.
- [27] N. Baurin, J.C. Mozziconacci, E. Arnoult, P. Chavatte, C. Marot and L. M.Allory, 2D QSAR Consensus Prediction for High–Throughput Virtual Screening. An Application to COX–2 Inhibition Modeling and Screening of the NCI Database, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 276–285.
- [28] P. Labute, A widely applicable set of descriptors, *J. Mol. Graph. Model.* **2000**, *18*, 464–477.
- [29] A. R. Katritzky, U.V. Maranm, S. Lobanov and M. Karlson, Perspective: Structurally diverse quantitative structure–property relationship correlations of technologically relevant physical properties, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1–18.
- [30] G. Moreau and P. Broto, The Autocorrelation of a Topological Structure: a new Molecular Descriptor, *Nouv. J. Chim.* **1980**, *4*, 359.
- [31] G. Moreau and P. Broto, Autocorrelation of Molecular Structures, Application to SAR Studies, *Nouv. J. Chim.* **1980**, *4*, 757.
- [32] Programs written by Gilles Moreau using mathematical library LAPACK
- [33] M. P. Gonzalez, J. Caballero, A. M. Helguera, M. Garriga, G. Gonzalez and M. Fernandez, 2D Autocorrelation Modelling of the inhibitory Activity of Cytokinin–derived Cyclin–dependent Kinase Inhibitors, *Bull. Math. Biol.* **2006**, *68*, 735–751
- [34] J. Caballero, M. Garriga and M. Fernandez, 2D Autocorrelation modeling of the negative inotropic activity of calcium entry blockers using Bayesian–regularized genetic neural networks, *Bioorg. Med. Chem.* **2006**, *14*, 3330–3340.
- [35] K. Fukunaga and W. L. G. Koontz, Application of Karhunen–Loeve expansion to feature selection and ordering, *IEEE T. Comput. C–19*, **1970**, 311–318.
- [36] Legendre regression program adaptation :
<http://www.bio.umontreal.ca/casgrain/fr/telecharger/index.html#Rgressionlinairemultiple>
- [37] mda program adaptation by Murtagh, <http://astro.u-strasbg.fr/~fmurtagh/mda-sw/>.
- [38] J. Gálvez, R. García, M.T. Salabert and R. Soler, Charge Indexes. New Topological Descriptors, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520–525.
- [39] E. Estrada, E. Uriarte, A. Montero, M. Teijeira, L. Santana and E. A. De Clercq, A Novel Approach for the Virtual Screening and Rational Design of Anticancer Compounds, *J. Med. Chem.* **2000**, *43*, 1975–1985.
- [40] E. Estrada and E. Uriarte, Recent Advances on the Role of Topological Indices in Drug Discovery Research, *Curr. Med. Chem.* **2001**, *8*, 1573–1588.

Biographies

Safia Tairi–Kellou is associate professor of theoretical chemistry at the University of Algiers. After obtaining doctorat she is engaged in teaching and research of theoretical chemistry and computational chemistry. In recent years, her main research interests are relationship between molecular structure and properties and she is collaborating on project related to computer aided Drug Design with Dr Bernard Maignet (CNRS, Nancy).

Souhila Bouaziz–Terrachet is postgraduate student at university of Algiers (USTHB). Her thesis under s. Tairi–Kellou’s supervision is a part of this paper.

Boubekour Maouche is professor of theoretical chemistry at the University of Algiers for the past 20 years. His major interests are theoretical modeling of reaction paths, including biochemical pathways, and determination of equilibrium structures of organometallic systems, especially in relation to biological systems.

Gilles Moreau is now retired. After studies in Ecole Polytechnique (Paris) and after obtaining a Ph D degree in organic chemistry at Faculté des Sciences d’Orsay (Paris–Sud University), G.M. joined a pharmaceutical company (Roussel–Uclaf, which merged in Hoechst and eventually disappeared). After leaving the company in 1997 he joined the modeling group at IECB (2 years, Bordeaux University) and then worked as consultant for various chemical or pharmaceutical start–up companies, up to 2005. His most preferred interest is for all aspects of chemical informatics and molecular modeling.